Plant science: Genomic compartments in barley

Keller, Beat ; Krattinger, Simon G

DOI: https://doi.org/10.1038/544424a

GENOMICS

Genome compartments in barley

Beat Keller and Simon G. Krattinger

Lead:
A high quality genome sequence of barley reveals a surprising compartmentalization of genes and repetitive sequences at the chromosome level and is a breakthrough for future genetic improvement of cereals.

Main text:
There are many ways to enjoy the produce made from members of the plant tribe Triticeae, for example in the form of freshly baked bread, perfectly cooked spaghetti or a cool beer. Triticeae play a pivotal role for our daily nutrition and include important cereal species such as barley, bread wheat, durum wheat and rye. An unpleasant feature of most Triticeae species however consists in their large genomes that even dwarf the human genome. These large genomes and the resulting lack of a Triticeae reference sequence posed a genuine problem for the improvement of these species because it made them unamenable to many modern molecular breeding approaches that are routinely used for other crops. Here, Mascher et al.[1] report the first complete and high-quality reference sequence of a Triticeae species by sequencing the barley genome. This genome sequence will not only impact future work in barley but cereal research and breeding in general. Notably the barley genome is also the largest genome of high quality available from any organism sequenced so far.

In ancient times, barley was a main food source in the Near East where this crop was also domesticated from wild relatives around 10,000 years ago[2]. Today, barley is still a primary cereal, best known for its importance in the production of beer and whisky but it also plays an important role as animal feed. Although genetic research on barley has accelerated in the last decade[3] no genome sequence of high quality (*i.e.* long contiguous stretches of DNA) was available.

The sequence assembly generated by Mascher and colleagues[1] was obtained by a hierarchical approach, starting with the sequencing of individual bacterial artificial chromosome (BAC) clones

of which each contained a tiny portion of the barley genome. These pieces were then assembled and extended to larger fragments by taking advantage of a previously established physical map[3]. In addition, a sophisticated method for sequence ordering based on physical proximity and distance-dependent decay of contact probability was successfully applied to linearly arrange and to orient the whole barley genome assembly. This approach has been demonstrated before for the human genome[4] and was successfully used now for a large plant genome. In total, 4.9 of the estimated 5.1 giga base pair (Gb) genome could be assembled and 95% of these assembled large fragments were linearly ordered, with half of all sequences belonging to DNA pieces of 1.9 Mb or longer. This allowed Mascher et al.[1] to develop seven chromosome-like sequences (pseudomolecules) matching the number of barley chromosomes.

A total of ~39,000 high confidence genes were annotated and positioned in the genome. Notably, the annotated genes only occupied 1.4% of the barley genome whereas more than 80% consisted of repetitive sequences. A genome-wide comparison of genome composition revealed in all chromosomes a striking pattern of three different compartments with distinct organizational and structural properties. There was an exponential increase of both genes as well as the ratio of genetic to physical distance towards the distal regions representing the ends of the chromosomes, while the proximal chromosomal regions showed lower gene content. The genetically dynamic distal compartments were enriched in rapidly evolving defense genes, while the proximal regions contained more housekeeping genes involved *e.g.* in photosynthesis and respiration. The repetitive part of the genome also revealed a highly uneven distribution of different repeat families along the chromosomes, but also in the close vicinity of genes. In the proximal chromosomal regions, there were distinct and ancient classes of elements, confirming the unique identity and evolutionary history of this part of the chromosome.

It has been known for some time that the proximal chromosomal regions, representing over 50% of the physical length of each barley chromosome, do not engage in novel combinations after crossing of two parents[3]. Therefore, the genes present in these regions are locked into large blocks that can hardly be broken up and reshuffled in the breeding process. The consequences of this are shown nicely in the study of Mascher et al.[1] for spring barley germplasm where the *HvCEN* gene, important for time to flowering and adaptation to geographical latitude, is mostly found in the same large block of identical genes. Obviously, breeders mostly selected for a specific form of the *HvCEN* gene ideal for spring barley, and all the other genes in the non-recombining region have been dragged along even if they might be unfavorable. Based on such knowledge on the gene complement in non-recombining blocks, future breeding efforts,

including gene editing or induced recombination, might result in highly desirable new gene combinations also in this part of the genome.

The work of Mascher et al.[1] is a breakthrough and game changer for barley genetics and breeding. However, this reference genome needs to be complemented by additional genomic information from many other barley genotypes. It is estimated that a single genotype of a specific plant species only contains ~85% of all the genes present in the total germplasm, the so-called pan-genome[5,6]. To make this diversity accessible for breeding, more genotypes must be sequenced at high quality. The partial sequencing of 96 elite barley lines by Mascher et al.[1] demonstrates serious erosion of genetic diversity depending on the genepool. This is essential and critical novel information to barley breeders for redesigning breeding strategies and for planning of pre-breeding activities.

The barley reference genome is not only valuable for breeding, but it will also give a boost to basic research questions in cereals. In barley, there are large collections of mutants which have been largely untapped because of the unhandiness of the barley genome. Such collections of mutants only now develop their full value as they become accessible for rapid gene identification[8,9]. Furthermore, comparative analysis with species such as durum wheat, bread wheat and rye will open exciting novel opportunities to better understand crop-specific biology and to study genome and chromosome evolution. Finally, an educated use of the barley genome knowledge combined with the recent establishment of CRISPR/Cas9 editing in barley[10] will allow sophisticated breeding interventions.

References

1. Mascher, M. *et al. Nature*, this issue (2017).
2. Riehl, S. *et al. Science* **341**, 65-67 (2013).
3. The International Barley Genome Sequencing Consortium *Nature* **491**, 711-717 (2012).
4. Burton, J.N. *et al. Nat Biotechnol* **31**, 1119-1125 (2013).
5. Hirsch, C.N. *et al. Plant Cell* **26**, 121-135 (2014).
6. The 1001 Genomes Consortium *Cell* **166**, 481-491 (2016).
8. Druka, A. *et al. Plant Physiol.* **155**, 617-627 (2011).
9. Sanchez-Martin, J. *et al. Genome Biol.* **17**, 221 (2016).
10. Lawrenson, T. *et al. Genome Biol.* **16**, 258 (2015).

The authors are at the Department of Plant and Microbial Biology of the University of Zurich, Switzerland.

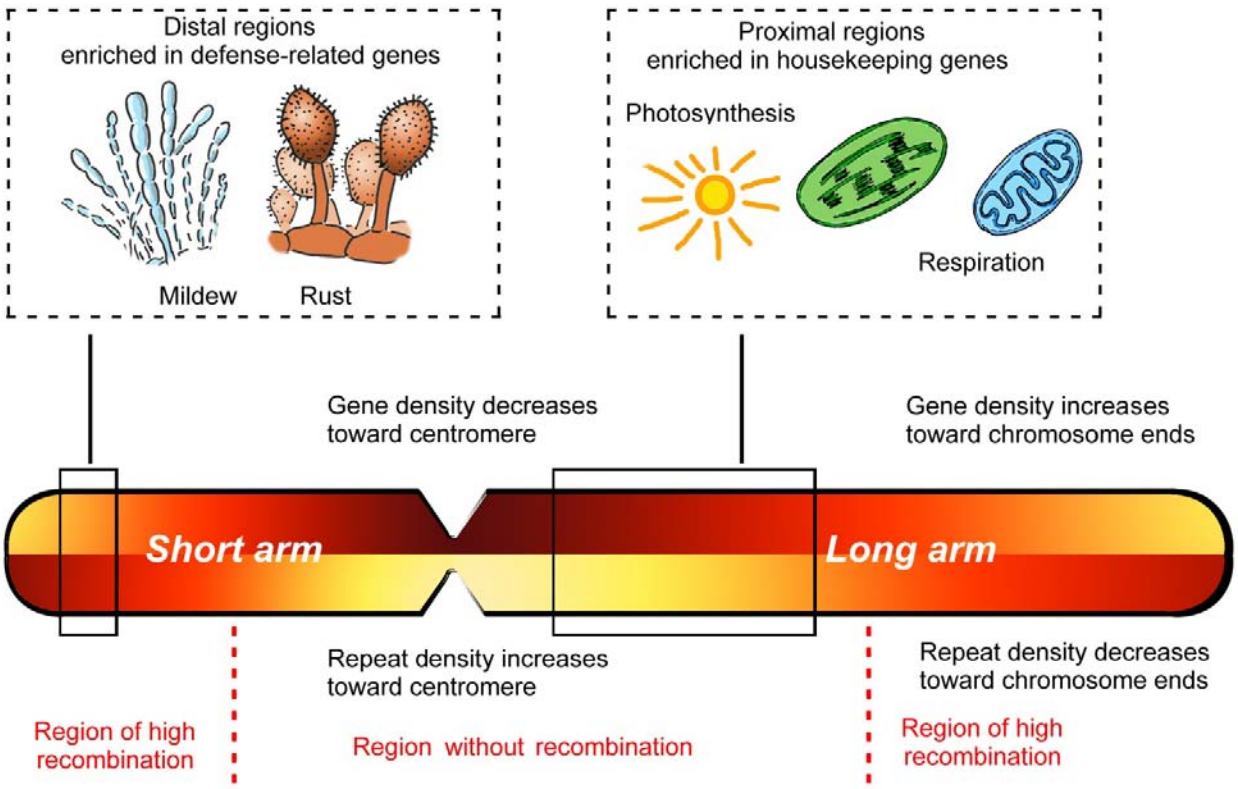

Figure 1. Schematic representation of a barley chromosome, indicating its compartmentalization. The distal, telomeric regions show an exponential increase of gene density while repeat density decreases. The telomeric ends are enriched in rapidly evolving genes (e.g. defense-related genes) while the proximal region contains more housekeeping genes (draft drawn by Dr. Thomas Wicker, University of Zurich, Switzerland).